

DOCUMENT RESUME

ED 205 582

TH 810 473

AUTHOR Kerins, C. Thomas
 TITLE The State Perspective on the Evaluation Burden.
 PUB DATE Apr 81
 NOTE 79p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).
 EDRS PRICE MF01/PC04 Plus Postage.
 DESCRIPTORS Bilingual Education; Compensatory Education; Federal Aid; Federal Programs; *Federal Regulation; *Federal State Relationship; Information Utilization; *Program Evaluation; Special Education; *State Departments of Education; Vocational Education
 IDENTIFIERS *Evaluation Problems; Evaluation Reports; Title I Evaluation and Reporting System

ABSTRACT

State and local education agencies increasingly have to spend a larger percentage of their time collecting, aggregating, and reporting information to federal agencies. This report suggests that the federal evaluation strategy to describe program implementation as well as impact is usually ill considered, inappropriate and of questionable utility. In some major programs more data are requested than is possible to supply; they are often asked for on an annual basis even though it is not analyzed annually; information demands are made in the absence of a plan for its use either for policy, program improvement or other analyses; and lack of coordination among federal agency staff requires duplication of effort and a further depletion of its time and fiscal resources. Finally, with few exceptions, federal evaluation efforts are unrealistic and inappropriate because of a tendency to avoid using the experience and expertise of local and state evaluation colleagues, an over-reliance on test scores, and a regrettable habit of pretending that categorically funded programs exist in isolation from each other and other important variables. As a result, federal evaluation requirements are viewed as a burden to be completed, not an opportunity for gaining a better understanding of a project.
 (Author/BW)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED205582

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

THE STATE PERSPECTIVE ON
THE EVALUATION BURDEN

C. Thomas Kerins

Illinois State Board of Education

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

T. Kerins

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Paper Presented at the American Education
Research Association Meeting. Los Angeles, CA

April, 1981

TM 810 473

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	1
INTRODUCTION.	1
FEDERAL ROLE IN EVALUATION.	2
CATEGORICAL EVALUATION.	3
TITLE I EVALUATION -- THE BACKGROUND.	6
TITLE I EVALUATION -- TWO APPROACHES.	8
VOCATIONAL EDUCATION.	23
SPECIAL EDUCATION	32
BILINGUAL EDUCATION	39
THE STATE EDUCATION AGENCY PERSPECTIVE.	45
RECOMMENDATIONS	50
BIBLIOGRAPHY.	63
APPENDIX A.	69
APPENDIX B.	73

EXECUTIVE SUMMARY

State and local education agency personnel increasingly have to spend a larger percentage of their time collecting, aggregating, and reporting information to federal agency staff. This report suggests that the federal evaluation "strategy" to describe program implementation as well as impact is usually ill considered, inappropriate and of questionable utility.

The requirements for information are conceptually legitimate as Congress passes laws to insure that the special needs of special children are met. This document describes examples of major programs in which more data is requested than is possible to supply; it is often asked for on an annual basis even though it is not analyzed annually; information demands are made in the absence of a plan for its use either for "policy", "program improvement" or other analyses; lack of coordination among federal agency staff requires duplication of effort and a further depletion of its time and fiscal resources. .

Finally, with few exceptions, federal evaluation efforts are unrealistic and inappropriate because of a tendency to avoid using the experience and expertise of local and state evaluation colleagues, an over reliance on test scores, and a regrettable habit of pretending that categorically funded programs exist in isolation from each other and other important variables. As a result, federal evaluation requirements are viewed as a burden to be completed, not an opportunity for gaining a better understanding of a project. Local and state personnel are forced into supplying independently derived numbers instead of developing a systematic approach to integrating evaluation efforts.

This paper's recommendations challenge local, state and federal personnel to collaboratively work together to streamline the data collection requirements, to redirect the resources presently allocated to the overcollection of questionable data to the training of those closest to the classroom about how to use information to improve programs for students.

Introduction

The Evaluation Team of NIE's Testing Assessment and Evaluation Division believed there was a need for an analysis of the impact of federal evaluation efforts on state education agencies. The term impact refers to problems and successes which have resulted from federally mandated evaluation/reporting systems as well as special issue evaluation studies.

Up until the mid-70's there was little concern about the need for more efficient and effective ways to use energy resources -- the supply was plentiful and inexpensive; only recently has our thinking refocused. There are a few members of the federal bureaucracy who seem to be on the verge of a similar discovery with regard to the collection of evaluation information from local and state agencies; there is a limit to the amount of time and care that individuals will take to accurately complete forms for a government that supplies 7% of the money and demands 60% of the paperwork (Allan, 1978). This situation is compounded when the image of the information movement is viewed as a consistent one-way flow to Washington.

One of the main reasons why this writer was chosen to author such an analysis is his association with the Council of Chief State School Officers' Committee on Evaluation and Information Systems (CEIS). The charge to CEIS is to provide the Chiefs with a single coordinated effort in formulating recommendations in the area of federal/state/local education information systems to reduce the data burden yet improve the quality of the data collected. As Chairman of the Evaluation Subcommittee from July 1977 to July 1980, this writer has had an opportunity to see first hand the increasing requests for evaluation data and the corresponding drain of time that

has occurred. While sensitive to this data burden, the author as well as the vast majority of his colleagues on the Evaluation Subcommittee are managers or staff members of state and local evaluation units and as such are certainly advocates for evaluation.

This paper then contains a review of the federal evaluation strategy from the perspective of professionals in evaluation who often have to produce and/or transmit data as well as analyses to the federal government and who are also potential consumers of these evaluations. While colleagues from the CEIS Evaluation Subcommittee have contributed to this paper, the responsibility for the final analysis and recommendations are solely those of the author.

Federal Role in Evaluation

The federal role in evaluation has expanded in its scope as the federal role in providing categorical funds for special children has increased. Either the collection of data for an evaluation or the submission of an evaluation report by local and/or state agencies has usually been required. In addition, the number of special evaluation studies sponsored by federal agencies has also increased. These evaluation efforts can be divided into two types. The first type -- The "Special Study" evaluation may focus on a particular aspect or objective of the program or it may provide a broad description of the program including what difference it made for the participants. The second type of evaluation requirement is the result of the establishment of a national system to collect statistics such as the number of participants, inservice training sessions held, parent advisory council meetings held, as well as impact information.

Stalford (1980), Pincus (1980), Kennedy (1980), Boruch (1980), J. David (1980), and House (1980) have all recently discussed the utility of these evaluation studies for local, state, and federal agencies as well as for the Congress which demands them. Through a review of this literature as well as other documents, and interviews with staff from the Department of Education, its contractors and legislative staff, this paper will particularly look at evaluation issues as they effect states. Examples in Bilingual, Title I, Special Education, and Vocational Education programs will be used to illustrate the issues, but first a brief discussion of categorical funding is necessary.

Categorical Evaluation

Berryman and Glennan (1980) cite the reliance on the experimental model by the federal government as the strategy for evaluating the large federal categorical programs. They argue that such a strategy inappropriately treats an education program as similar to a treatment in an experiment.

In large scale programs this approach implicitly assumes that children are discrete units of analysis that can be isolated, scaled, aggregated and reported. In practice a student in a Title I program may also be in a bilingual program as well as a vocational program and a special education learning disabilities class (all potentially funded jointly with federal state and local monies). A child could also be in a gifted class yet orthopedically handicapped and enrolled in a bilingual program. In addition, the

above children may be attending a magnet school funded through desegregation monies and influenced by an Operation "Push for Excellence" effort. Even with an experimental design, what regression or discriminant analysis can sort out which program had the greatest impact?

According to Weiler (1980), statements about cause and effect (Did categorical program x produce outcome y?) require the use of experimental or quasi-experimental research designs to rule out competing explanations for observed outcomes. Yet the conditions under which such designs can be implemented, and related conditions necessary for assessments of causality, are not present for studies of categorical education programs because:

- o Categorical programs are not educational practices, but ways of spending money accompanied by general resource-use guidelines. Programs do not have direct effects on children; only educational practices do.
- o Categorical programs are implemented in widely varying ways at sites that have very different characteristics, and key program elements (staffing patterns, management strategies, etc.) are not standardized across sites or over time. Most differences in outcomes are a result of these site and implementation differences, which swamp program influences.
- o Scientifically adequate control or comparison groups are ordinarily not available, and pre-program trends are not an adequate guide to outcomes that might have been expected in the absence of program implementation. Experimental designs require random assignment of subjects to control and treatment groups, and, for both groups, the elimination of all other influences (e.g., other programs) that could potentially affect relevant outcomes. Quasi-experimental approaches require the use of (nonrandomly selected) comparison groups, with the same treatment exposure guidelines. Because evaluations in complex natural settings cannot meet either of these requirements, there is no way to be sure of the extent to which a program is responsible for specific outcomes.

- o Criteria of program success are difficult to define and use, and the quality of evaluation instruments for measuring outcomes is largely inadequate. To escape an overly narrow focus on cognitive achievement, program guidelines sometimes specify board goals such as "responsiveness," "control," or "satisfaction." However, these goals are usually too broad to translate into precise evaluation criteria. (Weiler, 1980)

In interviews some National Institute of Education (NIE) officials frankly stated that the key problem is that despite the above critiques, the usual federal evaluation strategy is to look at what is happening from a categorical quasi-experimental perspective, not at individual children in their schools. Lois-ellen Datta stated that "we live in an era of single issue constituencies. While Congress and their staffs are aware of the need to know about the child as an individual they are personally accountable to special interest groups such as women, ethnic minorities, language minorities, the handicapped, and they therefore need data by category."

Legislative staff who were interviewed agreed with this analysis but also added the concept that Congress feels that many of its categorical programs are unpopular with local and state education agency staff because they are aimed at kids (i.e., migrant, handicapped) who otherwise would be ignored. Congressional assessment is that if federal legislation is not maintained in specific categories, services would decline; even if that did not happen, no one would have sufficient data to prove it one way or another.

Title I Evaluation -- The Background

Evaluation may be viewed as a government tool for policy analysis, accountability, or for program improvement at the federal, state or local level. The purpose, and therefore audiences, for any particular evaluation effort may be multiple or at least projected as multiple in order to be more palatable. Title I is an example of this.

Title I of ESEA in 1965 was the first social legislation to include a mandate for evaluation. The basic structure of the federal effort in Title I evaluation has remained the same for the last decade--a combination of special studies along with an annual system of local to state to federal reporting of local project effectiveness (TIERS--Title I Evaluation and Reporting System). A primary question for this paper is whether the time taken by local and state agencies to fulfill the data requests associated with these evaluation approaches has really served well any of the above purposes of evaluation.

The aim of evaluation can change over time just as the direction of a program can be modified. McLaughlin (1975) noted that the original purpose of evaluation of Title I programs was to provide parents the information they needed to ensure that their children were receiving the benefits of the new program. The purpose then of the evaluation was to hold schools accountable to parents. Subsequent guidelines issued by the USOE in the 60's added the notion that evaluation results should serve as a guide for improving the program (Program Guide #44).

Wargo (1972) analyzed the quality of the data produced by the TIERS system and concluded that state Title I reports were totally inadequate sources of information for assessing the national impact of Title I. McLaughlin (1974) noted that the mandated evaluation scheme had led to local evaluations that were typically seen as little more than an annual ritualistic defense of program activities. J. David's (1980) analysis was that: "Because the original legislation required that evaluation data be reported "up" from LEA's to SEA's and then to USOE, and because USOE could not produce usable aggregated data from the system, the dialogue between USOE and Congress consisted primarily of problems with the data that prohibited meaningful aggregation." David goes on to state that as a result there was never a serious discussion with Congressional staff as to whether such a system could ever produce the desired data. "Instead the problems with locally collected data, and the absence of any other available data set the stage for the way in which evaluation was viewed in the 1974 reauthorization of ESEA." House (1980) agrees: "Put simply, federal officials have attempted to evaluate these programs by an inappropriate approach to evaluation and when the approach fails, they modify the evaluation slightly and attempt the same approach again, and it fails again."

Congress was obviously frustrated by this in 1974. As a member of the staff of the House Education and Labor Committee put it: "Incredibly enough, both the sponsors and opponents found that almost no data existed on how much money from the Title I allocation was then being spent on basic skills" (Cross, 1979). Congress tried two solutions. One was to write into the

Title I legislation a requirement that the Office of Education develop "models of evaluation" and that local school districts evaluate their programs using these models. The other solution to the problem was to instruct the NIE to conduct a study of compensatory education.

Title I Evaluation -- Two Approaches

House (1978) maintained that to understand the evolution of the Title I models and their impact on state and local agencies, one needs to understand the driving philosophy behind most federal evaluations. He points to Rivlin's Systematic Thinking for Social Action as the key original document. In it she states that the key decisions in social services should be made at the higher levels of the federal government. The only way to gain sufficient knowledge about such services and programs is through experimental methods and statistical techniques. "Information necessary to improve the effectiveness of social services is impossible to obtain any other way" (Rivlin, p. 108). In using such techniques, according to House, it was possible to agree on goals and on a few outcome measures like test scores. Therefore programs should be organized to produce these narrow outcome measures for the federal analysts who would learn what was correct. It was not necessary to have experience or contact with the program in order to evaluate them; only strong centralized government could do the job. (House, 1978).

The Congressional Report documenting the reauthorization of ESEA in 1974 has a section which seems to support this perspective:

Another problem with evaluations is that there are no nationally uniform, and very few State-wide measures of evaluation which have been used in the program. Consequently, it simply is not possible to come up with one uniform set of national statistics, or with State-wide statistics within most states, to measure the effectiveness of the program.

The Committee believes that this lack of uniformity of evaluative practice has been a major failing of USOE. Federal administrators, many years ago, should have exercised decisive leadership in encouraging States to develop and require uniformity of evaluation among local school districts. (p. 4079 (US Code Congressional and Administrative News))

Within this same House Report though is contained a more pluralistic Congressional description of the purposes of Title I evaluation requirements:

The purpose of those (evaluation) requirements are: to enable each local educational agency to assess the effect of its program and to identify weaknesses as well as strengths of the project, thus serving as a tool for program revision and improvement; to enable each State to determine the extent to which progress has been made in reaching State goals for meeting the needs of educationally deprived children, as well as to provide a tool for State planning, management and dissemination; and to enable the Commissioner of Education to conduct a similar analysis at the national level. (p. 111).

As David notes, "For the first time, use of evaluation results by local districts to improve their programs is given at least equal importance to use of evaluation results for Federal accountability purposes." One might also add that the reference to State agencies using the results to assess State

goals was also a first for Title I legislation. The assumption was that the three-tiered (local to state to federal) reporting system could meet both accountability and program improvement needs at all three levels of program administration.

Section 151, an amendment concerned exclusively with evaluation was added to the reauthorization of the Title I bill. Funds not to exceed .5% of the program's appropriation were authorized to implement the evaluations as well as a mandated study of compensatory education to be completed by the National Institute of Education.

Section 151 required the Commissioner to "...provide to State educational agencies, models for evaluations of all programs conducted under this title...which shall include uniform procedures and criteria to be utilized by local educational agencies as well as by the State agency in the evaluation of such programs...the models developed by the Commissioner shall specify objective criteria which shall be utilized in the evaluation of all programs and shall outline techniques and methodology for producing data which are comparable on a statewide and nationwide basis." These models, using objective measurements, were to provide a basis for producing comparable aggregatable data at the state and national levels. Within five years USOE had prepared models of evaluations and regulations to mandate their use by local and state agencies.

Concurrent with these developments, USOE created a program for providing technical support and services to the states and LEAs. The major feature of this program was the availability of free technical services to SEAs and LEAs from 10 regional Technical Assistance Centers.

The Title I regulations are built around three "evaluation models" that were developed by Research Management Corporation (RMC) under a USOE contract. The models are purported to provide "as clear and unambiguous an answer as possible to the question, 'How much more did pupils learn by participating in the Title I project than they would have learned without it?'" (Tallmadge & Wood). Each of the three models is intended to be used to answer this question by comparing the observed test performance of students following participation in a Title I program with an estimate of what the performance would have been without the Title I project. The models differ in the way in which estimates are made of what the performance would have been without participation in a Title I project.

Is this what was meant by Congress? According to J. David's (1980, draft) interviews with Representative Veysey, the Congressional sponsor of the 1974 evaluation amendment, and the former Minority Staff Director of the House Committee on Education and Labor, Christopher Cross, there is some question. Veysey stated that if Congress had wanted Title I evaluated exclusively with standardized achievement tests it would have said so.

I don't exclude the idea of maybe some standardized tests where they can be made applicable or can be used. But that wasn't the total, or even maybe the major, part that we had in mind. We could have easily said that in very clear words...and we would have done

so if that's what we wanted, but that wasn't what we wanted. We had in mind a much broader thing than that, which could include test results where appropriate and could include a whole lot of other things that are not test results.

According to David's interview, Cross said that there were two themes of equal importance in Section 151: improving the quality of the data and building capacity at the local level. "To Cross, the word 'standards' meant giving locals and States 'acceptable procedures for conducting evaluations, uniformly acceptable practices across districts.' And 'models' referred to 'acceptable ways of analyzing and presenting the data.'"

In spite of some vague and inconsistent wording in the Congressional Report and in the language of Section 151 as well as the remarks of Congressman Veysey and staff, there was never any hesitation by the Office of Program Evaluation (formerly OED---the Office of Evaluation and Dissemination, formerly OPBE---the Office of Planning, Budget, and Evaluation) to proceed ahead with the national implementation of the models. The states were told to begin training workshops to prepare local districts for the time when data reporting through the models would be mandated via Rule and Regulation. The State Education Agencies were in an interesting dilemma. If the states stalled because of technical and other concerns with the implementation of these models to evaluate Title I programs, then the local districts could be required literally on an overnight basis to implement a complicated new assessment system while totally unprepared. On the other hand, if the SEAs hedged their bets and cooperated with the TACs in providing workshops

for LEAs in how to implement the system, they knew they would become another statistic, an example of how local and state staff warmly received the models and OPE's evaluation philosophy. Michigan is an example of the former, Illinois an example of the latter.

While renowned researchers such as Jaeger (1978) began the onslaught of procedural, technical, and other concerns about the models in the 1978 AERA convention, the CEIS Evaluation Subcommittee began expressing its concerns only a few weeks later at its annual meeting in Annapolis, Maryland. The catalyst for several years of heated discussion was the condescending manner in which some federal project officers informed state and local evaluation representatives of what they would be required to do and when. Unfortunately, this attitude hasn't changed dramatically. In a recent interview, the OPE liaison with the states said, "The SEAs and LEAs active in CEIS know the federal data collection efforts, but OPE doesn't know what SEAs and LEAs collect what data, nor the quality. Therefore, we can't even contemplate about whether we can use it. There exists in the minds of many federal employees a perception that SEAs and LEAs can't capably conduct evaluation."

The April 1978 CEIS Resolution (Appendix A) demonstrated that the Evaluation Subcommittee was simply not willing to accept the assurances of Federal officials that only minor and insignificant problems existed with the implementation of the system. Considering the tremendous costs in time and effort for local and state agency staff as well as taxpayer dollars that would have to be spent to implement and maintain the system, the resolution opposed the mandating of the testing models for Title I evaluation before there had been a careful review of the validity of the information that would be produced and its utility at all levels of government. Immediately

after the resolution was passed the director of OPBE sent a memorandum to the Chief State School Officers declaring that his office had completed activities which have culminated in "...the development of evaluation models that yield nationally comparable data." (Behind the CEIS resolution was the firm belief that OPBE had conducted initial informal reviews but, had not proven the above statement in a public and professional forum.) In this same memo to the Chiefs, however, it was declared that the models had undergone review "...by federal state and local officials and by experts in research methodology." The reply back to OPBE from the Evaluation subcommittee was that "If by review is meant that the models have been scrutinized in part by state and local education agency officials who have been attempting to accomodate the wishes of Congress, then the statement is accurate. If the term 'review' is meant to imply approval, then the statement is misleading." Despite personal appeals by OPBE personnel, the Chief State School Officers voted unanimously to support the resolution. The stage was then set for a truly unique confrontation.

After a flurry of letters between the Council of Chief State School Officers (CCSSO) and OPBE, a decision was made to hold a meeting in St. Louis during September, 1978. The Evaluation Subcommittee aimed for a meeting in which SEA and LEA evaluation directors would present their concerns about the models, the federal advocates would present their case, and a group of independent researchers agreed to by both sides would serve as a third force. This would all be done at federal expense since it was all part of the federal government's effort to prove the validity of the models.

Unfortunately, while money was overflowing to the 10 TACS, CEIS was informed that there were funds available only to pay for a few LEA or SEA expenses. In addition, OPBE felt that the technical advice for all parties could well be handled by their RMC consultants. So the States had to search out and pay for external research consultants.

At the beginning of the St. Louis meeting, the directors of evaluation in New Jersey, Michigan, Austin, Texas, and Philadelphia discussed their data based findings concerning the implementation of the models. OPBE did not respond but instead had a panel of Title I SEA directors declare that something was better than nothing and that the models would yield numbers that would fit into readable charts for state board review. (Subsequent meetings showed that both groups of SEA representatives had some of the same concerns, particularly about the ultimate use of the data. As a result, a much closer working relationship now exists among SEA program and evaluation staff in Title I and, more recently, in migrant education also.)

The external consultants did support many of the LEA and SEA concerns. For example, Linn (1978) assessed the models in this way:

Model B, the control group model, is the most scientifically rigorous of the three evaluation designs. Its technical merit however, depends on the similarity between treatment and control groups. Therefore, while Model B works well in the experimental laboratory the rigor is of no avail if the conditions of laboratory experimentation cannot be realized in the Title I evaluation setting.

As Linn (1978) pointed out, "Scientific rigor is illusory for the Title I evaluator who lacks the power to manipulate and control conditions and the power of assignment of people to conditions to be compared."

Model A is by far the most popular among LEAs and SEAs but is least preferred by its developers (Tallmadge and Wood, 1976) because it is apt to yield results that are seriously biased, a belief documented by the LEA and SEA representatives at the St. Louis meeting. Echternacht (1978) believed that the popularity of Model A might be expected largely due to its simplicity. It does not require complicated statistical manipulations, nor does it require that data be obtained on non-Title I children. "Evaluation requirements can be met in a relatively routine fashion only slightly more complicated than testing programs already in widespread use. Indeed, Model A is more appropriately identified as a testing model rather than an evaluation model. (Linn 1978).

During the dialogue at St. Louis and at subsequent meetings, two messages were given to OPBE by local and state agency evaluation personnel: First, there was not adequate support for the claim that when properly implemented each of the Title I evaluation models will yield valid, comparable, and interpretable results; second, that a much broader array of alternative approaches is needed to serve the various needs at the local, state and national level. As Linn (1978) pointed out:

Probably most important is the need to shift from an almost exclusive emphasis on student outcomes to a more balanced approach that promotes the evaluation of process and implementation. Achievement test results are an important component of a project designed, among other things, to improve student achievement, but they are only one, and often a relatively small part of what is needed for a good evaluation. Without knowledge of how, or if, a program is

implemented, test results are of little value. At the local level, descriptions of the instructional activities supported by Title I and the details of how the project was implemented are essential for the identification of project strengths and weaknesses since test scores alone do not provide indications of where change is needed.

The Title I evaluation models are designed to aggregate achievement test results to the national level. Since representative sample data would be sufficient for this purpose the investment involved in creating the models and their implementation must be justified primarily with reference to benefits at the local level (Barnes and Ginsburg, 1978). In 1978, David conducted a study designed to answer two major questions: do local staff use Title I evaluation results to identify strengths and weaknesses of their program in order to improve them; and are local uses of evaluation being changed because of the implementation of the model system? She found that, of course, standardized tests are used to meet federal requirements but she also found that they aren't used to judge the program or as a guide to program improvement unless skills-related tests, personal judgment and observation are incorporated in the decision process. Local staff also said that test results are often "not viewed as meaningful because important characteristics of schools (e.g., mobility) and children (e.g., socioeconomic status) have not been considered. Finally, in the eyes of staff and parents, the evaluation often excludes measurement of goals that they feel are as important as achievement" (David, 1978).

David followed up on the premise mentioned earlier. Since independent national studies provided the best source of evidence for the national impact of Title I, the most logical use of a Title I evaluation system would be to increase the usefulness of evaluation results for local and state per-

sonnel. While such a purpose is simply not a priority among Federal officials, until local and state evaluators view the federal endeavors in that light, effort devoted exclusively to the development of technically sound data will be quite limited, except perhaps to a few Washington bureaucrats. "The USOE evaluation models are designed to improve the quality of the data and will not, by themselves, lead to an increase in local use of evaluation. (David, 1978)

David (1978), Linn (1978), Wiley (1978) and Jaeger (1977, 78) all state that the use of any one test score as the cumulative index of project success is not accurate nor particularly helpful to anyone at any level of government either to make judgments or accurately describe impact. Yet enormous amounts of monies have been spent to develop and force-feed the models to do just that. Over 8.75 million dollars was spent just to fund the TACs for the first thirty months of their existence. The costs in time and dollars of local and state personnel are incalculable.

In the recent "Blue Ribbon" Committee evaluation report on the TACs, the issues of local use of evaluation and the aggregation concerns were discussed (Millman et al, 1979). "Expressions of disagreement with aggregation of data were stated strongly. This group of respondents included many TAC staff as well as SEA staff with greater than typical research training and some federal staff." In addition, "Almost all respondents believe that the major purpose of TIERS is to satisfy Congressional desires for summative information about Title I effectiveness. In fact, quite a few respondents 'complained' that this was the case, indicating a feeling that 'process evaluation' or 'formative evaluation' or evaluation 'of use to LEAs' would be a better purpose and do more good for Title I." Millman goes on to note

the contrast between "these perceptions of most SEAs, TACs, and some OED personnel" when compared to the perceptions of Congressional aides who were interviewed plus statements made in documents that were reviewed. These aides stated a belief that the Congressional leaders who originated the legislation for Title I evaluation were equally concerned about developing a system that would promote Title I evaluation at the SEA and LEA levels. "A Congressional belief was expressed that LEA evaluations would help to focus local attention toward needs for improvement in education." The "Blue Ribbon" Committee analysis corresponds to J. David's (1980) findings mentioned earlier.

This same report on the TACs noted that a majority of SEA staff interviewed did not have major reservations about implementing the models. These respondents felt that implementation of a model was "not a big deal" and were looking for any means to provide USOE with what it requires with minimum effort. However, the authors did not distinguish between the responses of SEA program and SEA evaluation personnel when they included that statement. Later on in their text they did state though that while "Most respondents are concerned about the accuracy of data from LEAs only a few of the SEAs are implementing any real quality control mechanisms."

In states where the minimum is the maximum expected, scant attention is paid to the quality of the data received. Since it is "no big deal", data are merely forwarded to Washington (where there is no known system for quality control.) In other states where the quality control issue is important, the frustrations mount, particularly when there is a high turnover rate among local teachers and administrators who complete the necessary reports. Some of these states are now asking for the raw data to do the scoring themselves

because of the high frequency of diverse errors. A major question here though is that even when states do attempt to centralize scoring to improve the mechanics of quality control, are they increasing accuracy at the expense of further reducing the meaning of the information and its potential utility at the local level?

Besides the "model" approach the other Congressional solution to the problem of inadequate Title I evaluation information was to directly instruct the National Institute of Education to conduct a study of compensatory education. The legislative mandate identified social problems for study, required NIE to submit an official research plan for Congressional review, established deadlines for interim and final reports to Congress, and provided a budget of \$15 million for the three year study. Finally, the mandate required NIE to submit its reports directly to Congress without any prior outside review. "By requiring reports in time for the next Title I reauthorization hearings, Congress showed that it wanted to use the results in legislative deliberations. By making Congress the sole judge of NIE's research plan, and the first recipient of its results, Congress stressed that its own interests were to dominate the study." (Hill 1980)

House (1980) in his assessment of the NIE effort notes that this was both an evaluation and a policy study since it assessed the past effectiveness of the program and examined specific alternative procedures. House goes on to note that NIE's wide focus enabled the study to avoid the floundering that occurs when standardized achievement scores are used as the exclusive indicators of program outcomes. "In both correlational and experimental designs, the Title I programs repeatedly failed to show large increases on standardized tests. Measurement and evaluation experts criticized the nar-

row focus of the outcomes, the experimental designs, the instruments employed, and the statistical techniques, thus throwing the evaluations into question. By contrast the NIE officials concluded that compensatory education had a number of important goals, most of which were not captured by outcomes on standardized achievement tests" (House, 1980).

NIE's 1976 interim report to Congress identified these goals for Title I:

- * To provide financial assistance to school districts in relation to their number of low-income children and within those school districts, to schools with the greatest numbers of low income students.
- * To fund special services for low-achieving children in the poorest schools.
- * To contribute to the cognitive, emotional, social, or physical development of participating students.

As NIE prepared its evaluation design, it is interesting to note the advice the project leader, Paul Hill, received from State and local personnel. "Researchers associated with state departments of education and local school districts were more concrete, urging us to conduct simple descriptive studies and to employ familiar and proven research methods. They recommended that we set modest goals especially for studies of the effects of Title I services on students achievement. These studies should acknowledge the diversity of instructional programs and judge the effectiveness of instruction only when its characteristics are well understood. The study should pay attention to the problems of day-to-day program implementation,

and try to understand the effects of federal and state regulations on local education practice" (Hill, 1980). Hill seems to have well understood too the diverse advice he would receive from different representatives of local and state agencies, noting they all have different internal divisions with different functions, goals, and points of view. (Hill's honest assessment is truly a rare occurrence since far too often federal bureaucrats will use a particular SEA staff member, such as a Title I director or Special Education manager for an endorsement and then claim that individual is speaking for the State or the States.)

The NIE study emphasized the fact that Title I is a program with many features, all of which had to be understood. It treated changes in student achievement as an important but not dominant topic. "By regarding student achievement as only one element of the program, we hoped to play it down as an evaluation criterion and to make the point that a program's effects on student achievement can derive from funding and administrative arrangements as well as from the quality of instruction" (Hill, 1980). It was the understanding then of the NIE project staff that simple measurement using available tests and models of gathering data and a common metric would supply information but would not answer the real questions of Congress---the real evaluators of federal programs. In fact, Hill believes that Congressional reluctance to act directly on the results of evaluations often occurs because the results are inadequate, perhaps too narrow or superficial, or too late.

In reacting to the NIE study the former minority counsel of the House Education and Labor Committee commented: "The NIE study has been remarkable in terms of the quality and quantity of information that it has produced. At this point one could say that the NIE study has been the single most productive piece of evaluation work ever undertaken on a federal education program." (Cross, 1979). In a recent interview another Congressional staff member stated that the timely completion of the reports was also a key factor, not only in the acceptance and use of the NIE Compensatory study but also the awarding of the Vocational Education Study to NIE.

In summary, the last five years have demonstrated two contrasting approaches toward the evaluation of compensatory education from the federal government. NIE produced a creative, thoughtful, useful series of studies in cooperation with a myriad of special interest groups, including state and local districts. On the other hand, OPE attempted to fix the leaky TIERS system regardless of the conceptual and technical problems raised by state and local evaluation personnel.

Vocational Education

The Education Amendments of 1976 (P.L. 94-482) were very prescriptive with regard to the evaluation of vocational education. State boards of vocational education in each state were directed to evaluate the effectiveness of all their programs assisted with funds under this act, and state advisory councils were directed to monitor the evaluations done by the state boards. USOE was mandated to conduct a review of all states' programs, and the National Advisory Council for Vocational Education was instructed to review the federal level administration and operation of vocational education pro-

grams including the reviews of the states' programs. Many other evaluation activities were stated or implied throughout the amendments as well as various planning, accountability and data collection provisions that were to be interrelated.

One mandated study directed NIE to "undertake a thorough evaluation and study of vocational education programs." The Minority Counsel to the House Committee on Education and Labor said that "The NIE study is a one-shot effort for congressional purposes, to get an independent evaluation of where vocational education is now." According to an NIE January 1979 planning report, "From the respondents at the federal, state and local level information will be obtained on the manner in which state and local education agencies implement federal policy for vocational education, particularly in regard to matters of compliance and evaluation." Specific information to be collected will involve answers to questions about the degree of state and local compliance, the manner in which state and local conditions affect the implementation of federal vocational education policy, the process of achieving compliance, the manner in which evaluation results affect program activity, the consequence of utilizing particular evaluation criteria, the impact of multiple federally mandated evaluation activities, and the nature of relationships between vocational education programs and those training activities operated through Comprehensive Employment and Training Act programs.

Why NIE? According to congressional staff it was because NIE had been responsive to congressional questions in the Compensatory Education study and had completed the task on a timely basis, two attributes not often emulated by other federal agencies with similar responsibilities. It is expected that NIE will give Congress what it wants -- a framework for discussion which will raise questions for focusing legislative amendments. NIE staff now working on this project also stated that as an institution NIE tends to assess the impact of legislation on the entire system -- all the consequences -- and not limit itself to just a single prespecified outcome such as a test score.

According to congressional staff one of the major reasons why this evaluation effort and other data requirements were included in the 1976 Amendments was the chaotic way in which data was collected and maintained by federal officials. In 1976, the available state vocational educational plans were neither compliance documents nor truly plans. In fact, those state plans available for inspection contained only the most recent amendment submitted by the state; the previous documents were not available so no one could inspect a comprehensive state plan. This is supported by the ECS review: "Prior to 1976, the Office of Education demanded a great deal of paperwork and detailed data from the states but there was no follow-up on this paper compliance" (ECS Report, 1979). (This example seems quite typical of the Washington malaise -- good intentions but rarely a practical plan to use the collected data.)

Throughout all the studies in this NIE endeavor, the relationship between CEIS and NIE was illustrative of how state evaluation personnel can work on a cooperative basis to both reduce the data burden and improve the overall project. NIE staff met on a continuous basis with CEIS representatives assigned to monitor all the studies as well as with the entire CEIS Evaluation Subcommittee. As noted in the NIE October 1978 progress report, NIE staff were careful in developing their plan for data collection to incorporate, as much as possible, data generated by other surveys or studies also called for in the Education Amendments of 1976. The dialogue with CEIS representatives was not solely to reduce the data burden but also to keep NIE staff and their contractors informed on the feasibility of collecting certain types of information. At one point in time an impasse was reached between CEIS and the contractor in the study concerning the distribution of vocational education funds. CEIS representatives agreed to a limited field test of the data collection instruments to prove to the contractor that the wrong questions were being asked. After the field test, the contractor admitted that it was impossible to collect the information according to his design and changed the instruments accordingly, thereby improving the eventual credibility of this particular study.

As per congressional requirement, on September 30, 1980, NIE project staff delivered their interim report to the House Subcommittee on Elementary, Secondary, and Vocational Education. The final report will be delivered exactly one year later. Among other areas, the report describes the effects of the 1976 Legislation upon evaluation. Those amendments required states to evaluate each of their federally-funded vocational education programs every five years and to assess those programs delivering entry-level job skills according to two criteria: students' placement in occupations related

to their training as well as employer satisfaction. The intentions underlying these requirements were that evaluation results would be used to manage and improve programs, ensuring greater congruence between vocational programs and labor market conditions.

One of the reasons for these requirements was that Congress discovered during hearings prior to the 1976 Amendments that funds were being distributed with little consideration of evaluation results, partly because of the lack of systematic programmatic evaluation. One of the major problems with evaluations as explained by Committee witnesses, was their failure to provide follow-up data showing job placement and satisfaction of program completers and leavers. A GAO report completed prior to the hearings stated that enrollments in programs supported by federal funds did not shift from traditional programs to programs for new and emerging job opportunities and attributed this in part to lack of systematic follow-up of students. In a recent interview with NIE project staff the example was used that Congress felt that without its intervention traditional but outdated printing programs would remain while laser technology would be ignored.

In order to bring about "rational" planning and to obtain information to judge the effectiveness of past congressional decisions a cluster of data requirements, including evaluation, was put into the law. "Congress seems to have had the best interests of vocational education at heart when members prepared these comprehensive amendments. They clearly felt that it was necessary not only to delineate evaluation requirements but also to try to ensure that evaluation data would be used for planning, improvement and redirection." (Datta, 1979) One major component of these requirements was the emphasis given to the employment success of program "completers" and

"leavers". According to the House Report, "The Committee has specified these two criteria to judge the effectiveness of programs because in our opinion, they show most clearly whether persons trained in vocational programs are showing the results of such training." A second key element of the evaluation provisions is that they be used to manage and improve programs, i.e., states were to conduct evaluations in order to assist local education agencies and other recipients of funds to operate the best possible programs of vocational education. However, neither the law nor the subsequent rules and regulations detailed how this was to be done.

State and local vocational educators objected to the use of the narrow criterion of job placements as the major evidence of success. They claimed that the goals of vocational education were broader than just training for jobs and that their programs should not be judged by whether the kids actually get jobs. Vocational programs should be evaluated on whether they have imparted the knowledge, skills, and competencies needed for those jobs available, but actual placement is determined by economic and personal choice factors beyond the school's control. There also was some concern that funds might eventually flow to those programs that could demonstrate "success" according to this criterion, thereby setting up convoluted games which might lead to the selection of easy to place students and the exclusion of harder to place disadvantaged students or women in non-traditional programs, the very groups Congress wanted to effect. (A similar concern has existed among Title I program personnel worried about the eventual use of a test score as the criterion for funding.)

USOE Regulations extended the requirements of the law substantially. Not only were state and local program evaluations required, but also there were four areas in which evaluations must be conducted. They included planning and operational processes, student achievement, student employment success, and the results of additional services. On one hand, the regulations expanded the criteria for evaluation so it wasn't narrowed to just student employment success but state administrators argued that they did not have the resources to implement all the evaluation requirements.

Nor did many state staff believe they had the resources to gather many of the data elements requested by federal officials as a result of Congress' desire for full and reliable information. NCES (National Center for Educational Statistics) was to set in place a national Vocational Education Data System (VEDS). Specifically VEDS was to provide information on vocational students (including information on their race and sex), programs, program "completers" and "leavers", staff, facilities and expenditures. Requiring all LEAs and SEAs to count students, programs, and staff in the same manner using standard terms was an effort to ensure accountability to Congress. In addition to providing accurate and reliable data for Congress, the system was conceived to be of utility to those planning and evaluating vocational education at the federal, state and local levels as well as supplying a basis for the comparative evaluation of vocational education with other programs for human resource development (e.g., CETA programs).

There has been an ongoing debate between CEIS and NCES since the inception of VEDS.

This dialogue has led to substantial reductions in the proposed system, but serious concerns remain that have prevented CEIS from recommending the system to the Council of Chief State School Officers. As we interpret VEDS, it would require the installation and operation of a student unit record system at the state or local level. Currently many state and local education agencies do not have recordkeeping and reporting systems designed that would provide the information at the level of detail mandated by VEDS. A significant effort in terms of time, human resources, and dollars would be required to permit all states to reach a comparable level of reporting capability. (Madden, 1980)

An earlier ECS report supports Madden's contentions. "The data requirements for VEDS imply both a student-based and an automated accounting system. While such a system may be valuable in the long run, initially it is difficult to implement in small schools or rural areas. Even states with sophisticated management information systems will have to make costly changes because of the requirements to use standardized definitions different from those used previously (ECS, 1979)."

A recent (May 1, 1980) survey to CEIS state coordinators, showed that the Enrollment/Termination data is still the most difficult to complete for most states with the employer follow-up requirements close behind. "Seventy-eight percent of the secondary and ninety-two percent of the post-secondary respondents believe that insufficient lead time for the SEA to develop or modify the computer systems is a major impediment. Also over 80% of the secondary respondents believe that insufficient lead time to prepare and orient LEA's to report the data will severely limit the quality of the Enrollment/Termination report." (Kerins, 1980) The problems which rise in meeting these federal requirements for follow-up data appear to stem not

from a disagreement with the concept of follow-up as one legitimate measure of effectiveness of vocational education, but from the procedural difficulty with gathering accurate and reliable data which are consistent across districts. The written responses to this survey indicated that the states which said they couldn't provide reliable and valid data in the detail requested were concerned with the issue of quality control and wanted to ensure that at least the best possible data would be forwarded. Personnel from other states indicated cynically that they would not exert any quality control review over the data submitted to them by LEAs because at the federal level there were no quality control mechanisms or clear data use plans.

The 1980 Interim NIE report notes that few if any states had in place in 1976 evaluation procedures as complex or as comprehensive as those envisioned by the legislation. "To date, the states have demonstrated a genuine commitment to and serious investment in evaluation, although they vary greatly in the scope and intensity of their efforts...the 1976 Amendments have strengthened the evaluation capabilities of the states" (H. David, 1980). However, some would say the federal government is pulling defeat from the jaws of victory by insisting on an impractical annual collection of student data -- a practice that is draining resources away from the intent of Congress. In addition, at a January, 1979 conference on Vocational Education evaluation, SEA personnel expressed concerns about the real and implied purposes of evaluation in the Vocational Education legislation. "Not only is there lack of clarity, but there is outright disagreement about the purposes of evaluation. And the purposes appear to be changing over

time. Evaluations that exist for program improvements, accountability, compliance, legislative input and/or as an impetus for change, are confusing. If purposes can be clearly stated and agreed upon, systems can be designed to achieve these outcomes."

Special Education

During the 1976-77 school year, the CEIS Evaluation Subcommittee and the Bureau for the Education of the Handicapped (BEH) now called OSE - Office of Special Education, had a continuous battle over an ambitious needs assessment project. If fully implemented as planned, every special education teacher and administrator in the country would have completed a 45 minute survey of very questionable quality concerning their "media and material needs". When it became evident that CEIS was not going to recommend approval, BEH officials attempted to bypass the organization and work directly with the State Directors of Special Education. Although the tactic did not work, it pointed out the need for each state CEIS coordinator to more closely communicate with his/her own SEA colleagues -- program managers who don't face on a daily basis the onslaught of one more survey or interview or form from an army of federal bureaucrats or their contractors.

The CEIS Evaluation Subcommittee realized that Public Law 94-142 would soon be in place and that numerous evaluation studies would be initiated. In an attempt to have SEAs and LEAs participate in the planning of the evaluation design, thereby potentially avoiding a repeat of the recent confrontations, the CEIS Evaluation Subcommittee passed a resolution (Appendix B) in April, 1977. The resolution, unanimously endorsed by the CCSSO, requested the BEH officials in charge of the special P.L. 94-142 evaluation studies meet with

representatives of CEIS (LEA and SEA evaluation personnel) during this planning process. The reaction of BEH leadership was much different than CEIS encountered one year later in the aforementioned Title I St. Louis meeting. In this case BEH paid all expenses for 25 LEA and SEA representatives to meet in Chicago during September, 1977.

During this meeting BEH presented five general principles as the basis for its P.L. 94-142 evaluation plan.

- 1) Evaluation activities should serve three audiences
 - a) the Congress via an annual report
 - b) USOE Program Monitors
 - c) SEAs and LEAs
- 2) Evaluation studies should be planned in close cooperation with SEA personnel.
- 3) The evaluation studies must cover the policy questions that the law explicitly or implicitly requires:
 - a) Are the intended beneficiaries being served?
 - b) In what settings are beneficiaries being served?
 - c) What services are being provided?
 - d) What administrative mechanisms are in place?
 - e) What are the consequences of implementing the act?
 - f) To what extent is the intent of the act being met?
- 4) The primary data resource for the BEH Evaluation Plan would be the information already being collected through other BEH data collection activities (State Program Plan, Child Count, End of year report and site visit data). Additional large-scale data collection is to be avoided. Supplementary evaluative information would be gathered through small focused efforts. These supplementary studies had four purposes:
 - a) To fill in gaps in information already being collected.
 - b) To shed light on how to interpret the national data
 - c) To identify relationships among variables already being gathered.
 - d) To identify or develop model practices.
5. Since the law allows for variability in practices selected by each state, the evaluation should document the variation rather than reduce the SEA and LEA activities to common elements.

The CEIS representatives were quite positive about the meeting. It provided an opportunity for SEA and LEA counsel on the federal 94-142 evaluation plans, thereby starting a healthy and mutually beneficial working relationship. The intentions of BEH to use existing data bases as much as possible and to concentrate on case study methodology in its special evaluation studies were welcome from the dual perspectives of data reduction and professional methodology. However, the CEIS representatives felt that there were some actions the Bureau should also consider since the only truly evaluative question was Question Five, the one concerned with the consequences of the law. BEH staff countered that while that may be true if other evaluators were the main audience, the primary audience was Congress and to them those were all evaluative questions of primary importance. (NIE's experiences in the Compensatory Education and Vocational Education studies would seem to support this viewpoint.) BEH staff also went on to state that once a data collection system was in place to routinely capture the information needed to answer the five compliance questions, a more qualitative approach could be implemented. The internal decision within USOE to keep the heart of any 94-142 evaluation studies within the Bureau and not with OPE corresponded to this initial approach of using evaluation as a formative tool and not to tie the new law down to premature summative evaluation efforts.

A second concern of the CEIS representatives was that only a vague ideal of intramural sharing and use of data within the Bureau was promulgated. Therefore, they requested the Bureau to:

1. Identify the data requirements for each of the six questions. (This should include an identification of the potential audience(s) for such data.)

2. Identify data sources at the federal, state and local levels. (This should include an analysis of the feasibility of data collection from each source.)
3. Define the means of gathering the data including methodologies, data collection systems and instrumentation. (This should include an estimate of the reporting burden for each of the data sources.)
4. Develop a data utilization plan including specification of analysis, verification, and feedback mechanism. (This should include an explanation of how data regarding the 11 handicapped categories will be analyzed.)
5. Determine the reporting methods to be followed, including an assessment of the feasibility of the aggregation of various types of data.
6. Determine evaluation reporting timelines and the necessary fiscal support for evaluation.

To some extent, progress has been made toward these requests over the last three years.

A third area of concern was the lack of utility of the evaluation findings at the local and state level. All the effort seemed once again to be channeled down a one-way street (local to state to federal) with little attention paid to their application in modifying SEA and LEA practices. The CEIS representatives requested that the Bureau "encourage and provide a means of rewarding those SEAs and LEAs which develop innovative evaluation procedures which exceed the minimum requirements of BEH... studies should be proposed that are suitable for bidding by state and local agencies as well as private contractors."

Three years have passed and where are we? First In 1979 OSE (formerly BEH) did demonstrate that it could merge data collected from multiple sources including some special evaluation studies, to produce the required congressional report - Progress Toward a Free Appropriate Public Education - on

schedule. However, according to a recent interview with OSE staff, there has been no official or unofficial reaction by Congress to this report. (Of note is the fact that the Human Resources Subcommittee Staff deals with all education programs and their evaluative reports such as the NIE Vocational and Compensatory Education studies, but does not handle OSE programs; that is done by another subcommittee which specializes in all areas of handicapped legislation). As of November 1, 1980, OSE has not yet sent the January, 1980 report to Congress.

Is evaluation a concern reflected in OSE policy and procedures? Quotes from Fall, 1980 interviews with OSE staff indicated not yet: "In 1980 Evaluation requirements are not now part of an SEA's annual plan since that plan serves basically as a compliance document. There is presently no system to review any ongoing LEA or SEA evaluation or to look across discretionary projects for evaluation data. OSE doesn't have statutory authority to require evaluations." However, these interviews also implied evaluation goals. "In the future, there will likely be specific evaluation requirements as we become more sophisticated beyond what present ordinary program review (monitoring) yields. We are in the compliance stage now; evaluation will follow. However, we still lack federal criteria/standards to use as evaluation models. Until that happens, how can we support statements such as 'That is a good program' or 'This a a bad program'?"

At the local level, the term evaluation refers primarily to diagnostic assessment of children. The regulations require that preplacement evaluation be conducted using multiple, appropriate assessment modes. If the child is found to have a handicapping condition, an Individualized Educational Plan (IEP) is devised. The content of the Individual Education Plan

is required by the regulations to include: 1) an assessment of present levels of educational performance; 2) a statement of annual goals and short term instructional objectives; 3) a statement of specific special education and related services and an assessment of the extent to which the child is able to participate in regular education programs; 4) projected dates for initiation and termination of services, 5) appropriate objective criteria, evaluation procedures and a schedule for reevaluation.

The state has a responsibility to ensure that the IEP has been prepared and that it meets the educational standards of the state. This is essentially a monitoring function and is carried out through on-site visits. Elaborate checklists have been developed by state agencies and BEH for assessing compliance with regulations. Additional monitoring requirements include fiscal audits and an assessment of the extent to which the Individual Educational Plan is actually carried out in practice. This latter function is essentially a check to ensure that the program for individual children is actually implemented.

The Secretary of Education has responsibility for evaluation under Section 618 of Public Law 94-142. Specifically, the legislation authorizes 1) annual studies; 2) assessment of the adequacy of information provided by state agencies; and 3) development of effective methods and procedures for evaluation.

Eleven special evaluation studies have been funded. They include such projects as a study of IEP implementation and erroneous classification of children. A recent OSE report on how these special studies improve program operation states that the evaluation studies will be used to do something to

LEA and SEA staff--improve monitoring, strengthen compliance, direct federal administrative action. Nothing is stated about local or state use of the findings to improve programs or procedures.

There are some noteworthy comparisons between the NIE Compensatory Education Study and the OSE Evaluation Studies. While the NIE staff was especially assembled for the task, had a direct relationship with Congress and conducted an evaluation of a program that had been underway a decade, OSE staff operated totally within the constraints of the Executive Branch and with a new program. Both NIE and OSE profited through their use of outside interest groups to form their broad range of substantive questions. According to Berryman (1980), the OSE studies are somewhat unique because they are part of a sequenced multi-year plan.

The CEIS Evaluation Subcommittee passed a resolution in April, 1980 when it became concerned about the "drift" of the evaluation studies effort in OSE, that is, a reduction in funds prohibiting new studies from being funded, three different directors in three years and a reduction in special studies evaluation staff. The resolution (Appendix B) stated that an appropriate time had passed since the 1977 meeting and that it would be timely to hold a national forum with LEA and SEA representatives to:

- 1) analyze how well special study evaluation efforts have been utilized,
- 2) ascertain the potential contributions of evaluation activities conducted by state and local education agencies to implement federal efforts, and
- 3) analyze the degree to which current data acquisition and evaluation activities meet the requirements of P.L. 94-142.

The Evaluation Subcommittee members hoped that among the outcomes of the meeting would be an increased utilization of evaluation findings at the local, state and federal levels, and general promotion of evaluation in the area of special education. The Chief State School Officers unanimously supported the resolution and added a potential outcome--that evaluation studies identify any modifications needed in the legislation. The goal had been for the meeting to take place by October 1980. Unfortunately, the forum will probably not take place until March, 1981 at the earliest. Why? In the response to the Chiefs' formal request the Assistant Secretary for OSE noted that the Secretary of Education has formed a task force on Equal Educational Opportunity for Handicapped Children to examine problems in four areas: data collection, enforcement, policy development, and technical assistance. "The Secretary's task force should lay the groundwork for all subsequent planning and development activities of OSE." The question then is presently unanswered as to whether the forum, along with the necessary pre-planning and follow-up activities, will be used as a pro forma session to review federal plans or as an opportunity for colleagues to assess the status and plan for the future development of evaluations as per the spirit of the CEIS Resolution.

Bilingual Education

The 1978 amendments to Title VII of the Elementary and Secondary Education Act required a thorough research and evaluation effort of bilingual education. In those amendments, Congress asked questions about a number of specific issues. Part C of Title VII requires coordination within the Education Department in order to answer questions such as who the children are who need bilingual education, how it may be used most effectively in class-

room instruction, and in what ways the federal programs can be shaped to support bilingual education. To coordinate the planning for an extensive research and evaluation agenda, a coordinating committee was established two years ago. The director of the Office of Bilingual Education leads this Part C Coordinating Committee. Other members include representatives from NIE, OPE, NCES and Policy/Planning Section of the new Department of Education. Also, Section 742 (c) of Title VII (Public Law 95-561) requires the Director of NIE to periodically consult with representatives of local and state educational agencies as well as other appropriate groups and organizations involved in bilingual education.

This Committee must oversee legislative mandates for studies and reports including Section 731 (d) the establishment of:

- 1) bilingual program models, including pupil-teacher ratios, teacher qualification, and other factors affecting the quality of instruction,
- 2) models for the evaluation of these programs in order to describe the progress made by participants in attaining English language skills.

Section 731 (e) follows on this latter point and further requires the Education Department to develop evaluation and data gathering models which take into account linguistic and cultural differences of the child in the context of existing state programs. Other data variables include "...pupil-teacher

ratios, teacher qualifications, length of the program, hours of instruction, percentage of children in the classroom who are English dominant and the percentage who have limited English proficiency."

In the USOE July 1979 Proposed Plan for Bilingual Education there is a brief section which elaborates slightly on the objectives of the evaluation models:

- 1) to identify the types of evaluation models most needed at the school district level,
- 2) to develop a set of models for use by LEAs,
- 3) to assess the need of local projects for technical assistance in project evaluation,
- 4) to pilot test the models in volunteer school districts. (optional)

In a recent interview with staff members from the Office of Bilingual Education (OBE) it was stated that a contractor is presently developing qualitative and quantitative approaches (models) for evaluating bilingual programs. The results of the contract will include handbooks, procedures, and forms but it will not include a requirement or even a suggestion for a common metric for reporting and aggregating scores. "These models will not resemble Title I." The understanding from program staff seems to be that the aggregation of student data across projects, districts, states and the country will not be important in the overall evaluation of Title VII.

Unlike Title I, Title VII LEA applicants receive direct grants from the Department of Education. The proposal review procedure specifies that each proposal is rated according to 110 points -- 15 of those allocated for an evaluation plan. According to recent interviews, evaluation information is submitted by forty-two states and six territories on 302,162 students for

almost 100 million dollars. Group data instead of individual student data is submitted directly to Washington by the approximately 500 local districts. Apparently the new models may change that to a student data base. Of note here is the fact that in the past the LEA project reports were stacked in boxes for storage. The claim is now made that descriptive data from these reports is being key punched and that a checklist is being used to review individual project evaluations. A sample of Title VII LEA project administrators interviewed for this report are quite disillusioned about the federal approach to evaluation up to this point in time. They rarely receive a reaction about their own evaluation -- whether it was at least adequate or even valid. They never know about other projects and their good or weak evaluations. There has been no standardized format as well as total inconsistency among their federal project officers. The staff of one project claimed that five different project officers had given them five different directions on how to complete the evaluation, when to submit it, and what would be acceptable. The term "analysis paralysis" was used to describe the federal approach to feedback on evaluation. Another director said that for the first three years he was told that "Evaluation means sending us anything you want." Then he was told to hire an external evaluator during the project's fourth year for a three day evaluation of the complete project--a waste of money in the eyes of the project director even though the external evaluator found that everything was fine with the project. A consistent plea was that someone in Washington who knows something about evaluation should be put in charge.

The Washington staff of OBE believes that there is light at the end of the tunnel. They admit past problems but claim that now OBE is hiring staff members who understand evaluation and research. As a result they believe that the image among LEAs of OBE with regard to evaluation will soon change. They have developed A Draft Guidebook -- Program Impact Evaluations as an introduction to evaluation for Managers of Title VII projects. The Guidebook doesn't spell out regulations but rather is designed to provide general information on the conduct of evaluations, evaluation methods, and ideas for evaluation of both the progress as well as the impact of Title VII projects. In the forward of this document it is stated that "This Guidebook is a draft and will remain in draft form for approximately one year. In the interim NIE and OBE, through the Title VII Network Centers, will continue to refine it."

OBE's Washington staff claim that they are very concerned about evaluation at the local level. "Making evaluation useful through documents such as the Guidebook is perhaps 50% of our effort with the rest aimed toward developing a system for federal data collection." They plan to present a standardized format this year when they meet with the Title VII LEA and SEA directors. Unfortunately, past meetings that have dealt with evaluations were described as not very useful at all by the LEA Title VII directors surveyed.

The law allows SEAs to apply for "technical assistance" contracts if, during the preceding year, an LEA within the state had received funds. These contracts may not exceed five percent of the total LEA awards. Activities associated with these contracts may take the following forms: 1) Monitoring

LEA Bilingual programs, 2) Evaluating the impact of programs, 3) Facilitating the exchange of information, and 4) Dissemination of materials acquired by the SEA to the LEAs. Boruch (1980) notes that:

While the regulations specify that the SEA application will be reviewed according to a point system and that the evaluation plan is to be specified, little guidance is provided as to what aspects should be considered. The review criteria for the evaluation plan is composed of a statement which simply mentions attributes such as quality of the evaluation plan, appropriateness of the methods and that methods should be objective and produce data that are quantifiable. Except for Basic Grants and Demonstration Projects, the same "boiler plate" statement appears in the description of the evaluation plan for all of the other program categories funded under this title. Such statements regarding evaluation components are of little use in guiding the development of an evaluation plan.

Unlike Title I, OBE does not collect its Title VII data through the state agencies. The states really play quite a minor role in assisting these local projects. However, most states do have an existing data system in place to obtain information about these projects since they often coexist with local and state funded bilingual programs. When asked if the proposed system of evaluation models and data collection will conflict with existing local and state requirements for evaluation, OBE staff frankly admitted that they didn't know. However, they stated they are seriously pursuing an analysis of present systems in order to avoid conflict as much as possible and to learn from the good systems that are in place.

There are already sufficient conflicts without adding to the difficulty. In one state the Office of Civil Rights (OCR) declared that a district had to use an oral language test as part of its screening procedures, but it could use any one. The district picked the "x" test but then the SEA said that

the district couldn't use that test and must instead use the "y" test. While the SEA only required that one of the four sections be administered, the "y" test only reports data on a composite basis. In addition, the Title VII project monitors gave conflicting suggestions. "Different directions from different sources on a continuous basis inhibit us at the LEA level from providing reliable data." The recently proposed "Lau" regulations may also serve as another layer of data. The August 15, 1980 Education Daily noted that "Identifying children eligible for proposed bilingual education programs could cost \$6 million to \$9 million, with more than \$100,000 of the cost tied up in paperwork, the government said. At least \$3 million would be in teacher time to distribute, collect and collate language data forms, the Education Department said in estimates of the cost of complying with bilingual education standards proposed earlier."

Through some of the activities mentioned above as well as others in the 1979 Proposed Plan, the Part C Research Committee is ensuring that all the pieces will be in place for the longitudinal study of the effectiveness of bilingual education to begin during the 1981-82 school year.

The State Education Agency Perspective

The degree to which SEAs are effected by various federal evaluation requirements such as the ones discussed in this paper vary exceedingly. As Boruch (1980) pointed out, some states easily surpass federal requirements and have developed remarkable evaluation reporting systems while other states make sure that they just do the minimum. When the individual Chief State School

Officer believes that evaluation is important in its own right and does not view it as another federal "lay-on", the resources are established for a strong SEA evaluation unit.

In a recent survey of SEAs by the Northwest Lab, Caulley (1978) points out a great diversity in the nature and size of their evaluative work load as well as the size of their staff (from one to 72). Some SEAs conduct all their own evaluations while others contract out, and still others do both, depending on the program, timelines, and resources available. Most states try to provide technical assistance so that LEA staff can better use the results of federal and state categorical program evaluation requirements, but there are a few that just "crunch" the numbers and send them to Washington. The evaluation philosophy which drives a particular SEA may be narrowly focused on student achievement through testing while other states may have a more selective strategy and use multiple measures, such as case studies, and adversary evaluation. Also, the evaluation approaches used will be affected by whether the SEA evaluation unit may be working with 24 regional offices containing qualified staff or over a 1000 LEAs, most of which have no one trained in evaluation. In either case, Boruch (1980) noted that within a given state, differences in the quality of LEA practices depend greatly upon state capabilities, interest and resources.

Boruch also stated that his visits to six states as part of his evaluation project caused him to believe that SEAs vary with respect to their level of involvement in the evaluation of federal programs. "Part of this state-to-state variation is due to the educational organization within a state, and some of the variation is due to the State's own investment in educational programs." For example, if a state has a strong financial interest in

bilingual education it is likely there will be a state legislative interest in evaluation which will serve as a driving force behind the development and maintenance of a strong state evaluation section. Boruch believes that the impact of state interest in evaluation on the LEAs practices can be remarkable, resulting in many LEAs substantially going beyond minimum evaluation requirements.

Boruch (1980) categorizes SEAs into three types according to their level of evaluation of federal programs. He tentatively characterizes SEAs as:

- a) Exemplary SEAs: This type of SEA has a well articulated evaluation plan encompassing both federal and state programs. The impact of this overall plan is seen throughout the SEA, across program areas and it permeates the LEA level influencing evaluation practices. There is strong support for evaluation from the Chief State School Officer and the legislature as well as strong public support for accountability.
- b) Emergent SEAs: This type of SEA no longer simply complies with federal regulations about evaluation. Instead, agency-initiated practices are developing with respect to how evaluations are to be carried out. In an exemplary SEA all or almost all of the programs are well evaluated whereas Boruch characterizes the emergent SEA as one in which good evaluations are not yet the rule but becoming more commonplace.

c) Compliance oriented SEAs: In this type of SEA inadequate staffing, monetary support and/or technical capacity (e.g., computer facilities) make it simply impossible to go beyond the minimal requirements established by federal regulations. Boruch informally makes a distinction here too between those SEAs in this category who make the effort to ensure that the compliance data is of high quality versus the ones which simply pass along data that is sent to them without a quality control check. The latter states seem to do that when they don't have the resources to complete the job or because they see that it is a meaningless exercise since no quality control efforts exist at the federal level to review elementary and secondary data, or because they really don't understand the ambiguous federal directions. An example of a combination of all these reasons occurred as recently as mid-October 1980. OSE and CEIS were not yet clear on how to count handicapped children who receive services under P.L. 94-142. Some states use the "snap-shot" one day count mandated by law in the fall as their end of year figure also, thereby avoiding a second count of students in June. Other states attempt a cumulative count of children who received services throughout the year. In some states this may mean the difference between reporting serving 7% or 10% of their student population. Nationally, no one knows what a final total means because some states submit the fall snapshot while other states send in what is theoretically an unduplicated end-of-the-year count of children who received services throughout the school year.

The majority of SEAs do have identifiable evaluation research and assessment units (Mathis). Boruch (1980) found in his telephone survey of 36 SEAs that 75% had distinct evaluation units or other offices which assumed at least

some evaluation responsibilities. Only in a minority of SEAs was evaluation totally located within program units. However, he also found that this responsibility varies. For example, no evaluation unit in Boruch's telephone survey had responsibility for all four major program areas discussed in this paper. "What this may suggest is the underutilization of evaluation units. While the infusion of federal monies and the associated evaluation requirements has generally facilitated the creation of these offices, this has not been true in every case...however, when there is a competently and adequately staffed evaluation unit, it can improve the quality of efforts, facilitate planning, and increase the use of information in decision-making" (Boruch, 1980).

Boruch's case studies of SEAs demonstrate that: 1) the SEA can play a variety of roles in the evaluation process associated with federal programs, ranging from compliance to sophisticated evaluation activity; 2) there are SEAs with the capabilities, interest, and administrative support to engage in high quality evaluation efforts. In fact, he also noted that in his teams' site visits:

...complaints were voiced by districts regarding the evaluation capabilities of SEA staff. These expressions of dissatisfaction, however, usually focused not on the quality of the services being currently provided by the SEA but rather on the additional services which the districts felt should be implemented. For example, large competent LEAs want the SEA to conduct special evaluation studies and issue waivers to districts so that they themselves can execute desired research. Title VII staff in local school districts want to receive technical assistance in evaluation on the same level as their Title I counterparts. Programs which employ outside contractors for evaluation want guidance as to the appropriate selection and monitoring practices in these arrangements. It should be noted, however, that the frequency and intensity of these criti-

cisms often paled against those which were leveled at the federal education agencies in terms of their lack of communication and feedback to the districts which provide them with evaluation data. (Boruch, 1980).

The lack of feedback is a consistent problem. In Fiscal Years 79 and 80 the Office of Program Evaluation funded a total of 26 grants to SEAs to refine the Title I Evaluation Reporting System. The grants were divided into the categories of Quality Control, Local Use of Evaluation, Measurement and Evaluation, Management Efficiency, Materials Development and Dissemination. Yet there has been no formal feedback or attempts to disseminate or evaluate the results of these studies except to utilize the TACs as libraries. Still statements are made from OPE that they don't know the capability of SEAs to conduct evaluations.

Recommendations

Kennedy (1980 draft) as well as others (Hall, Kerins, Chapman, 1980) point to the fact that each LEA has its own personality, its own atmosphere or climate which effects how staff view and use data. Boruch's case studies seem to generate the same image for SEAs. Therefore, in one sense specific statements about the SEA role in evaluation have to be considered in light of ever changing conditions among the 50 SEAs. On the other hand, there is no question that if evaluation practices are ever to live up to their potential in affecting decisions about education programs at all levels of government, then the role of the state needs to be changed immediately.

- o The States have to become an active partner in the planning of the national evaluation strategy.

On one level this means that it is not sufficient for congressional staff, NIE, OPE, NCES and others, including GAO (General Accounting Office) to meet and discuss what they will do to the SEAs. In Washington the Council of Chief State School Officers should have a voice in all these evaluation efforts from the planning stage. Without this change, many national evaluation efforts will continue to be controversial, confused and underutilized. With SEA involvement at least the opportunity for clarification and credibility should be increased.

Why? Reflect on recent events and planned actions in Title VII Bilingual Education as one example. The states with large bilingual programs of their own have either established or are establishing complex information systems. While there is at least an intention to review present SEA systems before imposing a national one, there is no doubt at this point, given past federal policy, which information system will have to change to accomodate the other, i.e., the VEDS experience. As a result, if a state has mandated a student I.D. system for establishing a student data base for annual and longitudinal data collection, the imposition of a federal system will either mean duplicate systems or the replacement of a tested system with the latest federal fad. While the example of the Title VII Research Committee is an excellent beginning in providing coordination among federal agencies, one notes that there is no state representative on this Committee. As the CEIS Evaluation Subcommittee Representative from a state with one of the largest

bilingual populations stated: "Washington officials act as if there is a total void but many SEAs have traveled these roads -- some good, some bad -- but at least we know the territory".

In another facet of bilingual education, the President of the National Association of State Boards of Education has said that recent proposed LAU rules do "...not take into account the strides more than a dozen states have already made in imposing sound bilingual education programs...and a handful of states as early as last spring offered to have their bilingual specialists work with Education Department officials to iron out trouble spots in the proposed regulations...our perception is the offers have been ignored." (Education Daily, 10/20/80). The quickest way for the Chiefs to portray an active as opposed to a reactive image is to put their commitment on the line in the planning stages.

Both Boruch (1980) and Hill (1980) point out the advantages, as well as some of the practical problems, of having Department of Education officials communicate regularly with congressional staff about evaluation questions and strategies. These advantages and problems also apply to representatives of the Chiefs--with the additional advantage that in many cases, the results of present state efforts can be brought to the fore in these discussions. Boruch (1980) recommends that "capabilities be assessed before new statutory evaluation requirements are directed at LEAs and SEAs to determine where resources are adequate to meet the demand...routinely consider pilot testing every new program, variations on existing programs, and program components before they are adopted at the national level, using high quality evaluation designs." An important opportunity was lost at the previously mentioned

St. Louis meeting because cities such as Austin, Texas and Philadelphia as well as the State of New Jersey had such Title I information ready for analysis but there was no formal mechanism to bring attention to the matter.

Another recommendation in Boruch's recent study was to "explore the feasibility and desirability of direct contract programs to capitalize on LEA and SEA capabilities." This happened recently with the contracts let in the TIERS refinement studies, but it is yet questionable if the "capabilities" were truly utilized. The recent CEIS resolution concerning P.L. 94-142 demonstrate the SEAs' and LEAs' concurrence with this recommendation. However, perhaps contractors are more easily controlled by federal project officers in their reports than would be peers in state and local districts.

Congress has hoped that many of the national evaluation studies and systems that have been developed and put in place would be of equal value to state and local educators. Also the recent years of debate over the paperwork burden placed on schools because of congressional mandates has made everyone more sensitive to the goal of more efficient data collection systems. Therefore, it seems logical that those individuals closest both to the physical work of collecting, maintaining, aggregating and reporting the data as well as attempting to use it for program improvement should somehow be involved in the overall development of the evaluation strategy. The national perspective must be maintained and coordinated by federal program officials in organizations such as NIE, OPE, and NCES but it needs to be tempered by those closer to the actual teacher/student classroom interaction than those individuals east of the Potomac.

- o The Council of Chief State School Officers (CCSSO) will have to seriously commit resources to staff a Washington office.

At the present time the CCSSO shares monies with NCES to jointly fund two professionals and a secretary in Washington. The mission of this project is to reduce the overall data burden from federal agencies by keeping each state CEIS representative informed of pending data collection and coordinating joint action when necessary. Each state annually contributes \$750 (maximum) to fund this project. A portion of staff time is spent with the Evaluation Subcommittee assisting the members track a myraid of federal evaluation efforts. If the prior recommendation in this section was cheerfully mandated by Congress and gleefully accepted by the Education Department, the Chiefs would be unable to maximize the opportunities for partnership under their present level of commitment.

This new component of the Washington office should not be dependent upon any federal funds. To accept any federal funding would jeopardize the credibility of recommendations and place the staff as well as the federal funding source in an awkward position if that agency also directs evaluation efforts.

This office would continue to work with the present voluntary efforts of the CEIS Evaluation Subcommittee members but would also attempt to broaden membership to bring more active participation from other state and LEA personnel representing both urban and rural schools. The expertise and experience exist in local and state agency personnel in the four categorical programs discussed in this paper as well as others. The key for the success of this concept is to provide the time and opportunity for the CCSSO staff to coordinate and pinpoint the resources and match them to the current issues. As

an example, this may mean pulling together perhaps five well respected SEA migrant program evaluators to participate in an ad hoc group that would conceptualize a national evaluation study not only to note what is needed but also what has been tried and what could mesh with existing systems.

Recently the Education Commission of the States' chairman, Florida Governor Bob Graham, stated that a vacuum exists among states "with regard to education policy", adding that it is being filled by the federal government through strings attached to federal money. In response to this recent (August, 1980) statement Education Secretary Shirley Hufstedler claimed the relationship between federal and state governments has too often been "adversarial...neither necessary nor productive," and should be improved. She went on to say that "Mistrust of state intentions seriously undermined federal incentives to coordinate with state activities, to take advantage of state resources, or to be sensitive to state concerns...if we are to achieve excellence at every level of American education, we must inaugurate an era of close state and federal cooperation."

- o Congressional evaluation requirements should be as explicit in intent as possible yet constrained by the knowledge that schools are not experimental sites.

The brief description of these four major categorical programs demonstrate that Congress is the main initiator of evaluation. How do the members perceive evaluation? Recently the Chairman of the Senate Committee on Labor and Human Resources said that "Evaluation holds promise of systematically and sensibly locating, collecting, analyzing, and interpreting the needed evidence. Yet the promise of program evaluation remains unfulfilled,

because we do not yet have the continuous information flow from evaluators essential to make policy and programmatic decisions of immense importance" (Williams, 1979).

Perhaps, generally the reasons why this sentiment is often also echoed by other legislators and their staffs is that there is an implicit assumption that all these categorical programs fit nicely together when placed in a local district just as if it were a giant jigsaw puzzle with neat interlocking pieces. Unfortunately, a more accurate image would be that of a canvas upon which colors have been indiscriminately splashed.

Federal program legislation usually requires a coalition of interests which typically leads to programs with ambiguous and multiple objectives. These federal education programs are delivered via states and localities that often have to reinterpret the programs according to their own unique situation including factors such as the SES of the community, characteristics of the student population, administrative procedures, union agreements, finances, and state legislative requirements.

Also, the statutory demands for evaluation are often ambiguous. As Boruch (1980) notes: The word evaluation can imply any activity from journalistic reporting to full-blown-field experiments dedicated to estimating the effects of an innovation on children. At worst, general demands to evaluate obscure the fact that feasibility of evaluation varies enormously and that elaborate evaluation may be unnecessary. Boruch goes on to recommend that Congress specify exactly which questions ought to be addressed and the appropriate audiences, whenever this specification is possible. Regardless

of how specific various requirements are in the law, he maintains that a persistent need for regular dialogue in refining questions and developing agreements should occur on a regular and not irregular basis as they do now.

When this dialogue occurs as it did in the two NIE studies there seems to be general satisfaction. Zweig (1979) wrote about the positive features of the NIE Compensatory study:

- 1) The policy-relevant questions were posed at the outset of the study in the legislation.
- 2) The study was designed as a multiple-component, integrated enterprise with the lead agency, the coordinator of a great many contractors of special analyses and evaluations.
- 3) Continuing consultation with congressional staff was built into the study's schedules.
- 4) Interim and final reports were geared to legislative events in the reauthorization cycle.
- 5) Reports were written in language everyone could understand.

As a contrast the reader has seen the problems that have occurred when Congress is not specific enough in its desires, as was the case with TIERS, and when it was too specific as with VEDS. The result was overkill in data collection/compliance reporting/evaluation requirements. Disagreement continues about the definition of the data elements, their aggregation, their utility, quality control, and the cost-effectiveness of the effort considering the data burden.

- o Redirect the current technical assistance strategy so that the implications of evaluation for policy can be accurately perceived by Congress and the potential for program improvement can be seen by people (parents and school staff) at the LEA level.

The view of how federal programs are implemented in local schools has been a linear (center to periphery) process (Farrar, 1980). That is, programs initiated by central (federal) officials were expected to be carried out at the periphery (LEA). The Rand Corporation's Change Agent Study (1978) saw motivation, commitment, and a sense of local ownership as important ingredients in program success. This meant there must be bilateral adjustment or mutual adaption between federal intentions and local wishes. Farrar maintains that both these views have shaped most evaluation studies of federal education programs. Federal program goals are the main criteria by which success is judged. With this perspective, evaluators have had to assume that federal programs are clearly specified. In reality this is seldom true. Most policies and programs are not templates, but bundles of potentialities or pre-dispositions waiting to be defined at the local level. (Majone, 1978).

Farrar goes on to propose a perspective with more explanatory power: Implementation is actually a complex and continuous process of policy evolution. These words imply that it would be fallacy to technicalize education as an activity by assuming most of the problems are "scientific" (Wiley, 1978). Lortie's (1975) claim that there is low interdependence among organizational units in schools and Weick's (1976) suggestion that schools are "loosely coupled" and do not always function in a highly coordinated manner are addi-

tional cautions. Yet often very prescriptive regulations define a federal evaluation strategy that pretends schools are suitable settings for experimental models.

Technical assistance, as exemplified in the Title I TACs, is constrained by this technical view of schools. The major problem with education evaluation, and the technical assistance designed to improve it is "...its reliance on the experimental model as a strategy for evaluating the large federal education programs. This model is most suitable for estimating program effects for small project grant programs" (Berryman and Glennan, 1980). Thus an inappropriate image of schools combined with an unsuitable evaluation model do not often present a strategy for evaluation that is particularly informative to consumers or to policy makers. "For the large federal education programs, in addition to knowledge of effects, policy makers need a broad range of descriptive information, interpretive information about program effects, and information about mature as well as new programs." (Pincus, 1980). Therefore, it seems logical that in the long run the responsible approach for federal technical assistance in evaluation is not to try to fit everyone to preconceived Washington molds, but rather to ensure that sufficient resources are available for LEAs and SEAs to conduct high quality evaluations that meet their needs. Future national studies that would tap into these systems, perhaps on a case study basis, would find more valid information.

These thoughts are hardly new. The "Blue Ribbon" Committee that evaluated the TACs recommended that the Education Department investigate alternative systems, perhaps including a more pluralistic model, that would supplant the present systems. "On the one hand, it would be surprising to find that the

technical assistance center approach has been entirely abandoned in the TAC Program of 1983. On the other hand, it would be equally surprising to find that the TAC approach has not evolved by then into a more flexible and probably more complex system". (Millman, 1979).

The goal for Congress, Education Department officials, as well as SEAs should be to maintain standards of credibility yet slash the amount of data requested and the frequency of requests made to local districts.

An hour after a presidential debate a good sampling firm can determine the attitudes of America with only a plus/minus four percentage point difference. Yet many bureaucrats and legislators apparently cannot see that for many categorical programs such sampling procedures could also apply, whether for compliance or evaluation purposes.

The TIERS and VEDS systems demand that vast resources be spent in the collection of minute data about all the participants instead of building local expertise so that valid evaluations could be the rule and not the exception. The choice is between a continuing emphasis on building systems that demand a continuous flow of aggregatable data to a Washington bureaucracy that simply can't digest it, or structuring a system that will enable a reasonable amount of information to be gathered at reasonable frequencies.

What does the future look like? According to a recent interview with the coordinator of OPE, "Federal data requirements will grow in the future, requiring more complex evaluation and monitoring information systems. Data

must be comparable, aggregatable; the standardization of information that is occurring in Title I is a key because it points to how future data burdens can be more efficiently handled."

In its FY 80 Plan, NIE stated its belief that a major concern of the federal government should be to improve the quality of local school district programs through evaluations. "The way in which evaluation is conceptualized at the federal, state, and local levels, however, is not necessary conducive to program improvement at the local level." The authors of the NIE Plan note that monitoring and accountability functions are equated with the broader concept of evaluation, and as a result evaluations are not used directly as a means of improving programs despite congressional intents. This view is shared by many CEIS Evaluation Subcommittee members who believe that once prescriptive regulations are put in place initiative is restricted because LEA and SEA resources are then aimed solely toward compliance.

The results of national special study evaluations have not usually been well utilized at the policy level in Washington. Some recent interviewees suggested that federal program officials couldn't react to evaluations because of the nature of their roles. They exist to administer money and to check compliance. However, OPE is seeking to dispel that image. When a project is complete, the OPE project officer chairs an ad-hoc group that reviews the final report and develops option papers which could effect the program's management, budget, or legislation. These option papers are then recommended to the Secretary who will make the final decision on whether to adopt, adapt or reject the policy options. This ad-hoc group would then have some implementation oversight responsibilities. This system is part of a strategy to speed up the clearance process and to see that the evaluation

results are used. Unfortunately, no CEIS or LEA/SEA representatives, the ones likely to be most affected, will be members of these ad-hoc groups unless the first recommendation in this report is followed.

Bibliography

Allan, John. "The View from the States: Perspective on Federal Forms Clearance Procedures." Paper delivered at the American Education Research Association meeting, Toronto, Canada, 1978.

Barnes, R. and Ginsburg, A. "RMC Models - Their Relevance for Title I Policy Concerns." Paper presented at the meeting of the American Educational Research Association, Toronto, March, 1978.

Berryman, Sue and Glennan, Thomas. "An Improved Strategy for Evaluating Federal Programs in Education" in John Pincus (Ed.) Educational Evaluation in the Public Policy Setting. The Rand Corporation: Santa Monica, California, 1980.

Boruch, Robert and Cordray, David. An Appraisal of Educational Program Evaluations: Federal, State, and Local Agencies. Northwestern University: Evanston, Illinois, 1980.

Caulley, Darrel, & Smith, Nick. Field Assessment Survey. Northwest Regional Educational Laboratory: Portland, Oregon, 1978.

Cross, Chris. Title I Evaluation -- "A Case Study in Congressional Frustration." Educational Evaluation and Policy Analysis. 1, p. 15-22, 1979.

Datta, Lois-ellin. "Better Luck This Time: From Federal Legislation to Practice in Evaluating Vocational Education" in Theodore Abramson (Ed.) Handbook of Vocational Education Evaluation. Sage Publications: Beverly Hills, 1979.

David, Henry. Vocational Education Study: The Interim Report. National Institute of Education: September, 1980.

David, Jane. Local Uses of Title I Evaluations. Research Report for the Office of the Assistant Secretary for Planning and Evaluation. Washington, D.C., 1978.

David, Jane. Federally Mandated Evaluation Policy: The Evolution of the Title I Evaluation Program (draft). Bay Area Research Group: Palo Alto, California, 1980.

Echternacht, Gary. "The Use of Different Models in the ESEA Title I Evaluation System." Paper presented at the meeting of the American Educational Research Association, Toronto, March, 1978.

Evaluation of Vocational Education: Roles, Responsibilities, and Responses of State and Federal Agencies. Education Commission of the States: Denver, Colorado, 1979.

Farrar, Eleanor, DeSanctis, John, and Cohen, David. "The Lawn Party: The Evolution of Federal Programs in Local Settings." Phi Delta Kappan: Bloomington, Indiana, November, 1980.

Hall, Melvin, Kerins, Thomas, and Chapman, Carmen. Student Evaluation Practices in Selected Illinois Districts. Report submitted to the Illinois State Board of Education, January, 1980.

Hill, Paul. "Evaluating Education Programs for Federal Policymakers; Lessons from the NIE Compensatory Education Study" in John Pincus (Ed.) Education Evaluation in the Public Policy Setting. The Rand Corporation: Santa Monica, California, 1980.

House, Ernest. "The Objectivity, Fairness, and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation." Educational Evaluation and Policy Analysis, 1, p. 28-42, 1979.

House, Ernest. Evaluating for Public Policy: A Review of the National Institute of Education Compensatory Education Study. Center for Instructional Research and Curriculum Evaluation: Urbana, Illinois, 1980.

Jaeger, Richard. On combining achievement test data through NCE scaled scores. Research Triangle Institute: Research Triangle Park, N.C., 1977.

Jaeger, Richard. "The Effect of Test Selection on Title I Project Impact." Paper presented at the meeting of the American Educational Research Association, Toronto, March, 1978.

Kennedy, Mary. The Role of Information in Public School Management and Instruction (Draft). The Huron Institute: Cambridge, Massachusetts, 1980.

Kerins, Thomas. Mimeo Report presented to CEIS Executive Board. April, 1980.

Linn, Robert. "Comments on the Proposed Regulations and Models for Title I Evaluations". Paper written for the Illinois State Board of Education, Urbana, Illinois, 1978.

Lortie, Daniel. School Teacher: A Sociological Study. University of Chicago Press: Chicago, 1975.

Madden, Vincent. Letter to the Honorable Mark Shedd, Commissioner of Education, Connecticut, March, 1980.

Majone, G. and Wilkavsky, A. "Implementation as Evolution: Exorcising the Ghosts in the Implementation Machine." Russell Sage Discussion Papers, no. 2, Sage Foundation: New York, 1978.

McLaughlin, Milbray. Evaluation and Reform. Ballinger Publishing Company: Cambridge, Massachusetts, 1975.

Millman, Jason, and others. Performance Review of UOSE's ESEA Title I Evaluation Technical Assistance Program. Hope Associates: Washington, D.C., 1979.

Pincus, John (Ed.) Educational Evaluation in the Public Policy Setting. The Rand Corporation: Santa Monica, California, 1980.

Rivlin, Alice. Systematic Thinking for Social Action. The Brookings Institution: Washington, D.C., 1971.

Stalford, Charles. Evaluation Team Research Program--FY 80. Testing Assessment, and Evaluation Division. National Institute of Education.

Tallmadge, G. and Wood, C. User's Guide: ESEA Title I Evaluation and Reporting System. RMC Research Corporation: Mountain View, California, 1976.

Wargo, M.J. ESEA Title I: A Reanalysis and Synthesis of Evaluation Data from Fiscal Years 1965 through 1970. American Institute for Research: Palo Alto, California, 1972.

Weick, Karl. "Educational Organizations as Loosely Coupled Systems." Administration Science Quarterly, March, 1976.

Weiler, Daniel and Stearns, Marian. "The Uses and Limits of Education Evaluation at the State Level" in John Pincus (Ed.) Educational Evaluation in the Public Policy Setting. The Rand Corporation: Santa Monica, California, 1980.

Wiley, David. "Evaluation by Aggregation: Social and Methodological Biases." CEMREL, INC., Chicago, Illinois, 1978.

Williams, Harrison. Foreward to the book Evaluation in Legislation edited by Franklin Zweig. Sage: Beverly Hills, California, 1979.

Zweig, Franklin. "The Evaluation Worldview of Congressional Staff" in
Franklin Zweig (Ed.) Evaluation in Legislation Sage: Beverly Hills,
California, 1979.

APPENDIX "A"
CEIS RESOLUTIONS
CONCERNING
THE TITLE I EVALUATION AND REPORTING SYSTEM (TIERS)

CEIS RESOLUTION -- April 28, 1978

WHEREAS, Section 151 of the Education Amendments of 1974 requires the U.S. Commissioner of Education to provide models to state education agencies for evaluations of all programs conducted under Title I which shall include uniform procedures and criteria to be utilized by local educational agencies and shall result in comparable data nationwide;

WHEREAS, since 1974 the U.S. Office of Education has developed such models and has contracted with Technical Assistance Centers to advocate, explain and assist state and local education agencies in the implementation of the models;

WHEREAS, the models have been implemented in some school districts and states on a voluntary basis and have raised the level of interest, awareness, and commitment to sound evaluation of Title I;

WHEREAS, technical data from use of these models is only now becoming available;

WHEREAS, these data and special studies of the models raise questions about the validity, reliability and utility of the models;

WHEREAS, additional time and resources are needed to evaluate existing models;

WHEREAS, any attempt to mandate the models through rule and regulation, program guidelines or otherwise may have the effect of mandating models which will yield inaccurate and misleading data and,

WHEREAS, CCSSO supports the collection of reliable and valid data for program evaluation;

NOW THEREFORE BE IT RESOLVED that CCSSO request the Secretary of HEW to delay any plans to mandate the use of specified evaluation models until such technical and methodological problems as are now arising are resolved.

CEIS RESOLUTION -- October 26, 1978

WHEREAS, USOE, in response to requirements of Section 151 of the Education Amendments of 1974, is developing a set of models for evaluation of ESEA Title I;

WHEREAS, staff of some state and local education agencies, and institutions of higher education, have raised serious technical questions regarding the models and their applicability to the Title I population;

WHEREAS, CCSSO has expressed its opposition to requiring the use of the models until such time as the emerging technical problems can be resolved;

WHEREAS, the proposed regulations and models reflect only a testing/reporting system and not a comprehensive method of evaluation needed to provide Congress or the decision-makers at federal, state or local levels with the information needed to make sound decisions regarding program changes that will facilitate program improvement;

WHEREAS, the proposed regulations will force some local and state education agencies that already have effective evaluation systems to replace them with a system that will provide less information or add to their current system at substantial additional costs;

WHEREAS, USOE has indicated that approximately 50% of the nation's Title I districts are currently implementing one or more of the models; and

WHEREAS, this number may be sufficiently large to enable USOE to allow an objective evaluation of the models;

NOW THEREFORE BE IT RESOLVED the CCSO urge the Commissioner to:

- (1) encourage the continuation of
 - a) the utilization of USOE Title I models where SEAs and LEAs find them appropriate and meaningful for their purposes, and
 - b) technical assistance programs as an integral part of improving the quality and utility of evaluations;
- (2) present for wide distribution and discussion a plan for the continuing analysis of the validity, reliability and utility of the models, including a methodology for aggregating and/or reporting data obtained by alternative evaluation practices;
- (3) publish the regulations with a three year time limit
 - a) indicating the models are developmental and need further analysis, and
 - b) amend the proposed language to eliminate the constraints imposed on the development, implementation and use of alternative models, and
 - c) allowing SEAs and LEAs the option of continuing their present evaluation systems during this 3 year period.

CEIS RESOLUTION -- April 24, 1980

WHEREAS, CEIS views the evaluation function at all levels--local, state, and national--as critical to impacting on educational issues and for providing educational data for decision-making as well as policy setting and policy resolution; and

WHEREAS, CEIS supports the need to provide evaluation data which are gathered and aggregated in a systematic and uniform process on Title I programs; and

WHEREAS, Public Law 95-561, Section 183, requires the development, adoption, and implementation of a uniform and systematic evaluation reporting system for Title I; and

WHEREAS, CEIS has previously expressed concern over the mandated Title I reporting system through a resolution of October, 1978; and

WHEREAS, concerns continue to be expressed regarding methodological, implementation, and technical problems with the Title I Evaluation and Reporting System (TIERS); and

WHEREAS, similar questions are expressed regarding the further refinements of the Title I Evaluation and Reporting System (TIERS); and

WHEREAS, an evaluation study of the Title I Evaluation and Reporting System is being planned;

NOW, THEREFORE, BE IT RESOLVED that the Council of Chief State School Officers request the U.S. Department of Education (1) to immediately accelerate the planning of an evaluation of the Title I Evaluation and Reporting System (TIERS), and (2) to ensure that the evaluation is implemented in an unbiased fashion.

APPENDIX "B"

CEIS RESOLUTIONS

CONCERNING THE EVALUATION OF PUBLIC LAW 94-142

CEIS Resolution -- April 21, 1977

WHEREAS, Public Law 94-142 was signed into law on November 29, 1975; and

WHEREAS, Public Law 94-142, Section 618(A) states "The Commissioner shall measure and evaluate the impact of the program authorized under this part and the effectiveness of state efforts to assure the free appropriate public education for all handicapped children"; and

WHEREAS, Public Law 94-142 Section 618(b) states "The Commissioner shall conduct directly or by grant or contract, such studies, investigations, and evaluations as are necessary to assure effective implementation of this part..."; and

WHEREAS, the proposed Rules and Regulations published on December 30, 1976, pursuant to Public Law 94-142 state "...that each annual program plan must include procedures for the evaluation at least annually of the effectiveness of programs in meeting the educational needs of handicapped children..." (Section 121(a) 47); and

WHEREAS, the proposed Rules and Regulations published on December 30, 1976 pursuant to Public Law 94-142 state that within 90 days after the end of the grant period, the state education agency shall submit a report to the Commissioner on the activities carried out under this part during that period; the report must contain the results of the evaluation and in brief narrative form the impact these funds have had on the states educational services to handicapped children ages 3-5; and

WHEREAS, there has been little attention to the statutory requirements regarding evaluation while the Rules and Regulations and Annual Program Plan have been formulated by the Commissioner's staff; and

WHEREAS, the data requirements for state evaluation and federal evaluation should be as timely, clear and consistent as possible, and

WHEREAS, the CCSSO believes the implementation of program evaluation requirements are of great concern,

NOW THEREFORE BE IT RESOLVED that CCSSO request the U.S. Commissioner of Education to provide an appropriate forum no later than July, 1977 to define common purposes and discuss certain program evaluation procedures and requirements for local and state education agencies as required by P.L. 94-142; and that the CCSSO request the U.S. Commissioner include invitations to federal, state and local personnel.

CEIS RESOLUTION - April 24, 1980

WHEREAS, Section 618 of the Education for All Handicapped Act of 1975 (P.L. 94-142) mandates data acquisition and evaluation activities; and

WHEREAS, the initial period of implementation and evaluation is coming to a close; and

WHEREAS, CEIS and representatives of B.E.H. jointly met in September, 1977 when the original P.L. 94-142 evaluation questions were first discussed;

NOW, THEREFORE, BE IT RESOLVED that the Council of Chief State School Officers request the Assistant Secretary for Special Education and Rehabilitation Services to sponsor an appropriate forum, not later than October 1980, to review the data acquisition and evaluation activities in light of the requirements of P.L. 94-142, and to invite federal, state, and local education personnel, including CEIS Evaluation and Data Acquisition Subcommittee members. The forum should include:

- (1) An analysis of the degree to which current data acquisition and evaluation activities meet the requirements of the Act;
- (2) How evaluative findings, particularly special study evaluation findings, have been or will be utilized;
- (3) The potential contributions of evaluation activities conducted by state and local education agencies to complement federal efforts; and
- (4) Identifying further implications for influencing federal legislation.

The outcomes of this forum would be to:

- (1) increase utilization of evaluation findings at the local and state level;
- (2) impact on future directions for evaluation and data collection activities under P.L. 94-142;
- (3) promote evaluation in the area of special education; and
- (4) influence subsequent legislation.

RWD:1597f